LBNL, NERSC and ESnet Update

Juan Meza
Department Head
High Performance Computing Research
Lawrence Berkeley National Laboratory
http://hpcrd.lbl.gov/~meza

Office of Science
U.S. DEPARTMENT OF ENERGY

NERSC

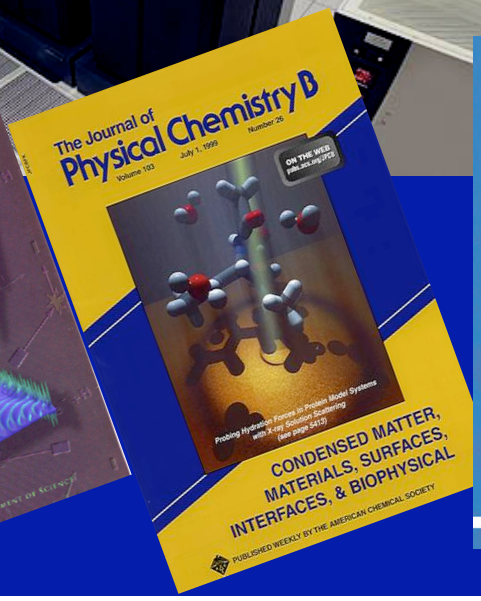
ESnet
Energy Sciences Network

National Energy Research Scientific Computing Center

Serves the entire scientific community

- ~2500 users on ~250 projects
- Focus on large-scale computing

NERSC



NERSC Center Overview

- Funded by DOE to support open, unclassified basic research
- Delivers a complete environment (computing, storage, visualization, networking, grid services, cyber security)
- Focuses on intellectual services to enable computational science
- Provides close collaborations between universities and NERSC in computer science and computational science

Large-Scale Computing Is Critical for the Success of SC Programs

- Increased computational modeling in base program
- Increased reliance on NERSC resources by major user facilities, e.g., RHIC, ATLAS, JGI
- Success of SciDAC and INCITE, and need for their follow-on programs
- New programs and facilities include computation as critical element, e.g., GTL, fusion simulation initiative (ITER), nanoscience, JDEM

NERSC Strategy

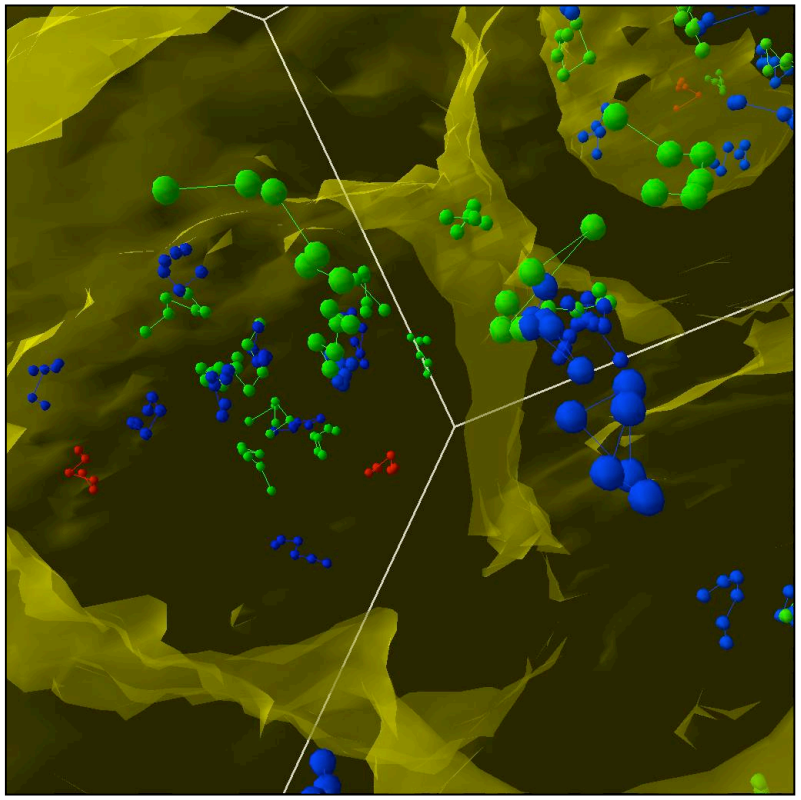
- Deliver large scale, leading edge, balanced high performance technologies for computation, storage and networking in an open and high performance manner.
- Provide systems and services to meet the diverse computational science need of DOE and the nation including major initiatives such as Grand Challenge, SciDAC and INCITE.
- Provide comprehensive support for capability computing for all ranges of scientific users through consulting, special processing, algorithmic support and collaborations.
- Foster close interactions with the computational science users through frequent interactions, formal and informal feedback and clearly understanding the user's requirements.
- Are leaders in Science Driven Computer Architecture by closely collaborating with the science community, leading industry suppliers and other large scale computing facilities to create new and improved technology and methods for scientific computing.

Impact of NERSC for Science Mission

Majority of great science in SC is done with medium- to large-scale resources - In 2003, NERSC users alone reported the publication of at least 2,404 papers that were partly based on work done at NERSC.

INCITE: Quantum Monte Carlo Study of Photosynthesis

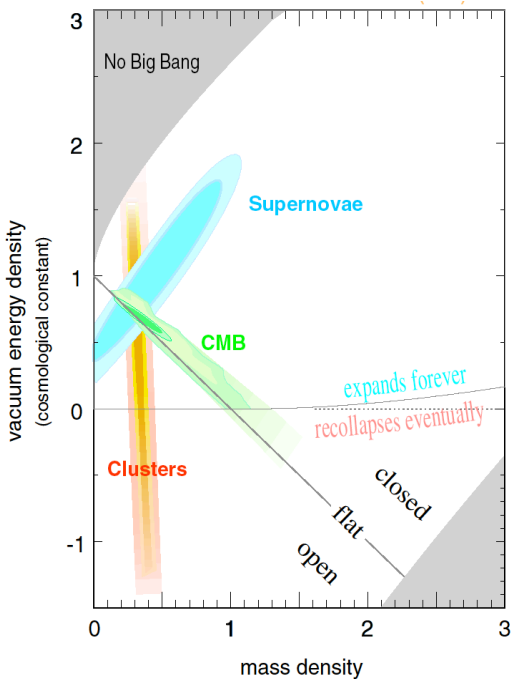
- PI: William Lester and Graham Fleming, LBNL/UC Berkeley
- Goal: determine the ground to triplet-state energy difference of carotenoids present in photosynthesis
- Computation: Zori code for diffusion Quantum Monte Carlo, scaled to 4096 processors
- Results: most accurate values of the excitation and total energies of these biologically important systems; largest QMC calculation ever

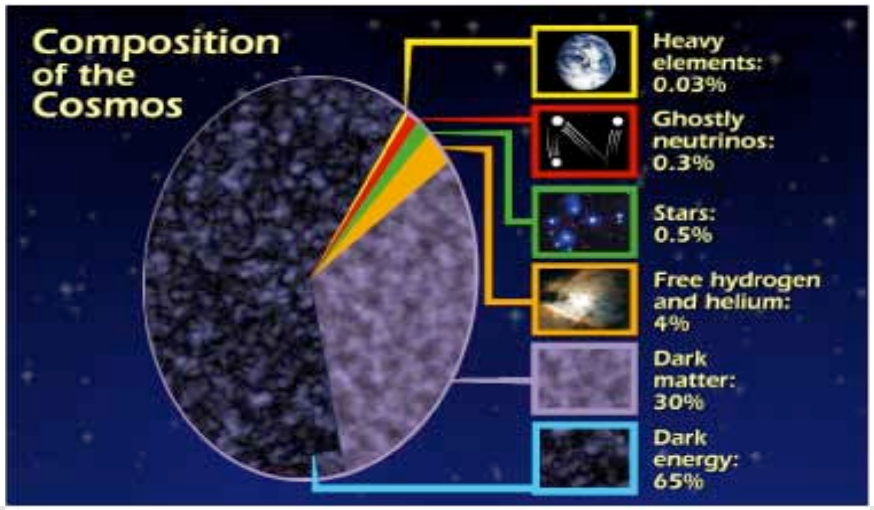


Imaginary time paths traversed by electrons in a photosynthetic system. The electrons are colored to make them distinct. The yellow isosurface shows the boundary of the molecular framework.

College of Chemistry
University of California, Berkeley

Experimental Cosmology

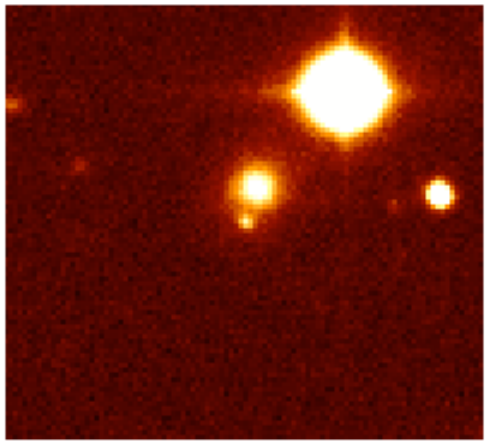


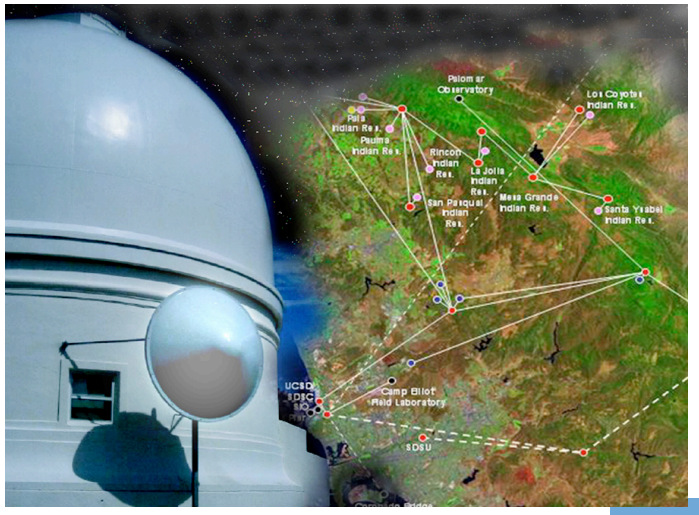


- Methodology for early detection of type Ia supernovae - Saul Perlmutter
- Cosmic Microwave Background (CMB) radiation – Julian Borrill
- Most distant supernova detected – Peter Nugent
- Largest set of distant type Ia supernovae detailed with Hubble telescope - Saul Perlmutter, Greg Aldering, Rob Knop, et al.

Nearby Supernova Factory

- Goal: Find and examine in detail up to 300 nearby Type Ia supernovae
  - More detailed sample against which older, distant supernovae can be compared
- Discovered 34 supernovae during first year of operation and now discovering 8-9 per month
- First year: processed 250,000 images, archived
  - 6 TB of compressed data
- This discovery rate is made possible by:
  - high-speed data link
  - custom data pipeline software
  - NERSC's ability to store and process 50 gigabytes of data every night



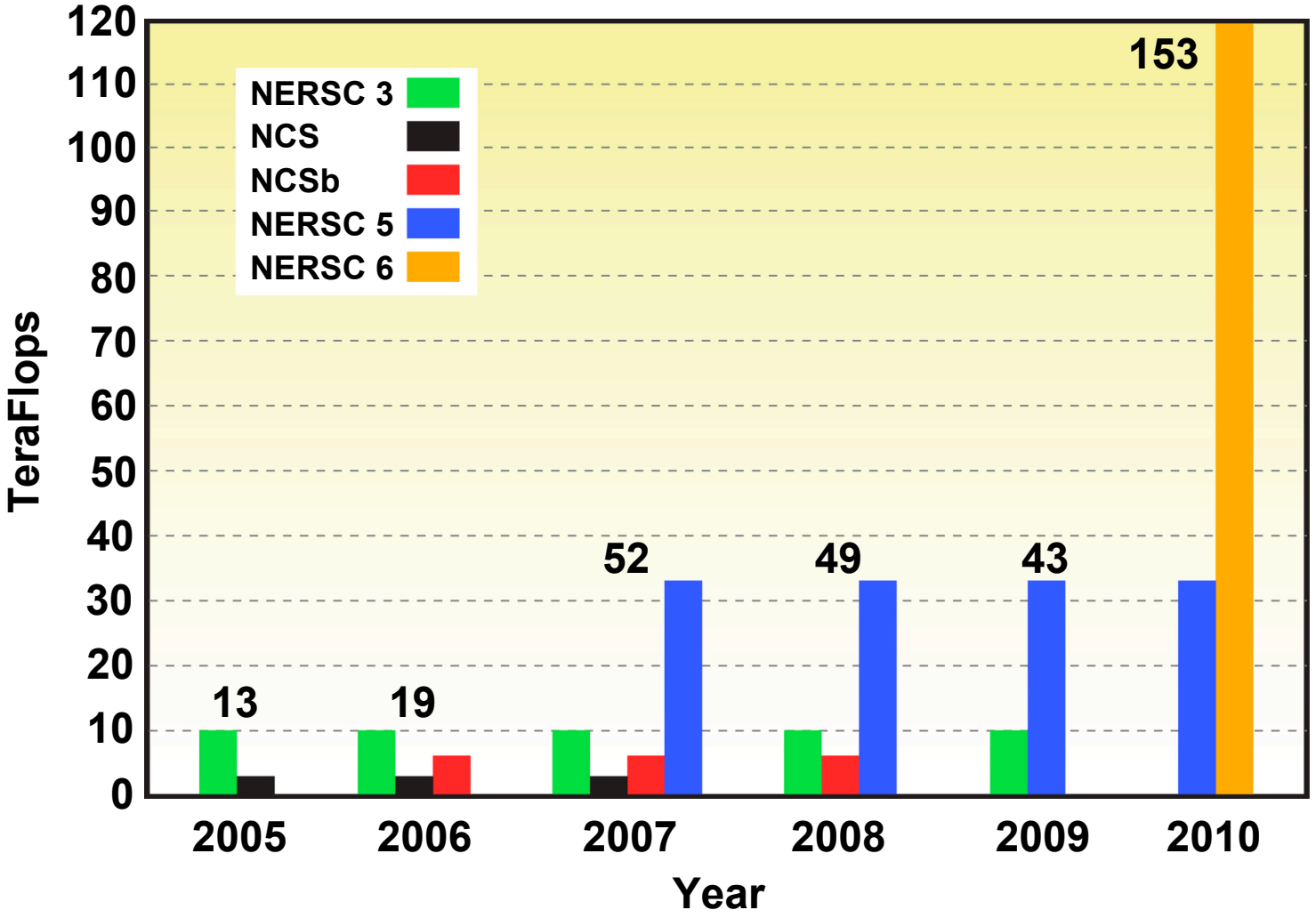


NERSC Baseline Plan, FY05 – FY10

NERSC will have several production systems at the same time.
Two major systems and multiple smaller systems.

- **NERSC 5: 2007 initial delivery**
  - **3 to 4 times Seaborg in delivered performance**
  - **~33 Tflop/s peak, 6.6 Tflop/s sustained**
- **NERSC 6: 2010 initial delivery**
  - **3 to 4 times NERSC 5 in delivered performance**
  - **~120 Tflop/s peak, 30 Tflop/s sustained**
- **NCS and NCS-b**
  - **Interim, focused systems**
  - **NCS 2005 – about 30% of Seaborg**
  - **NCS-b 2006 – about 60% of Seaborg**
- **PDSF will continue to double every year in processing power/disk**
- **HPSS and network will scale in proportion to computational systems**
- **Servers, visualization, grid support, cyber security**

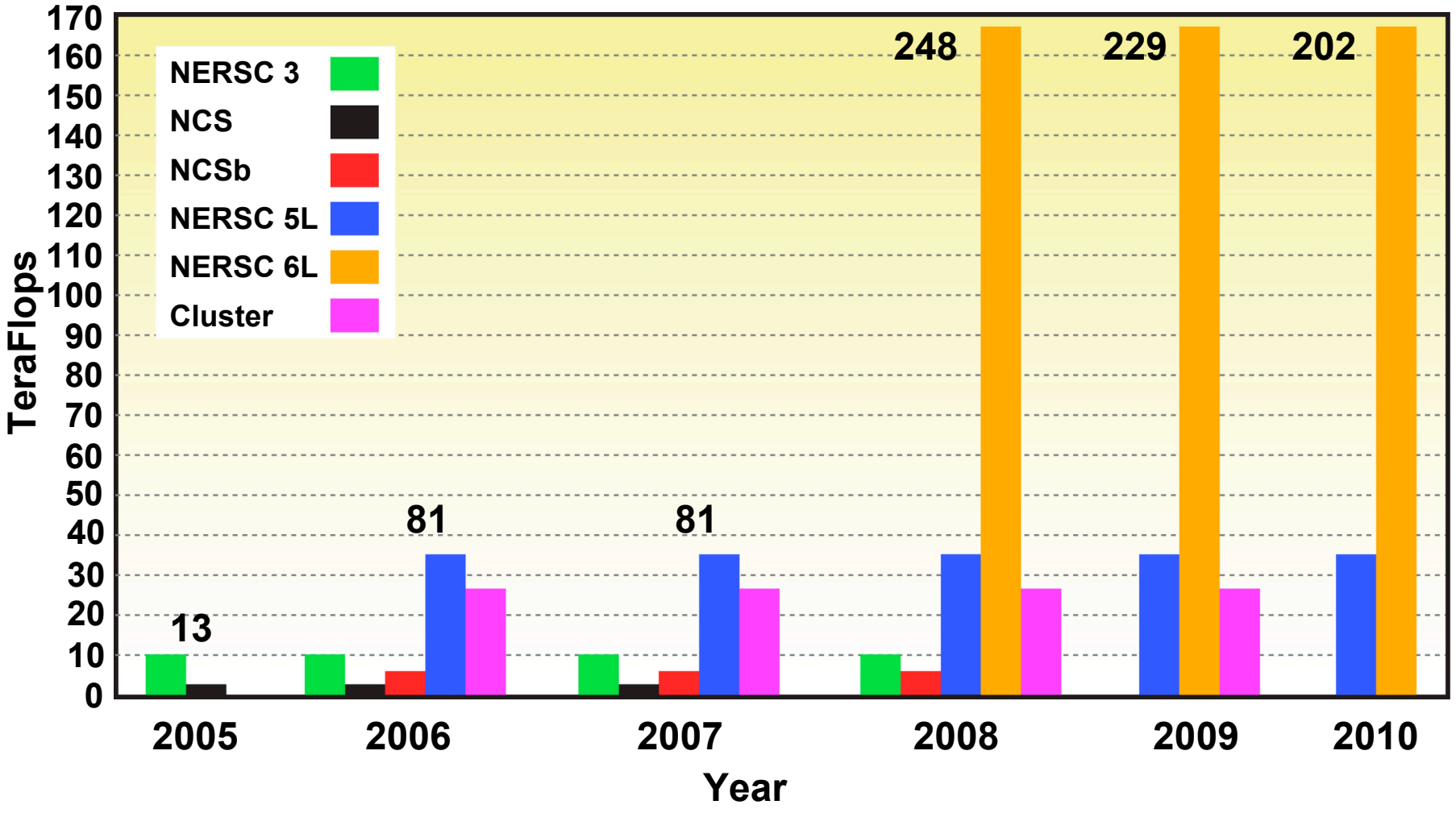
NERSC Baseline Plan



Detailed NERSC Capability Plan

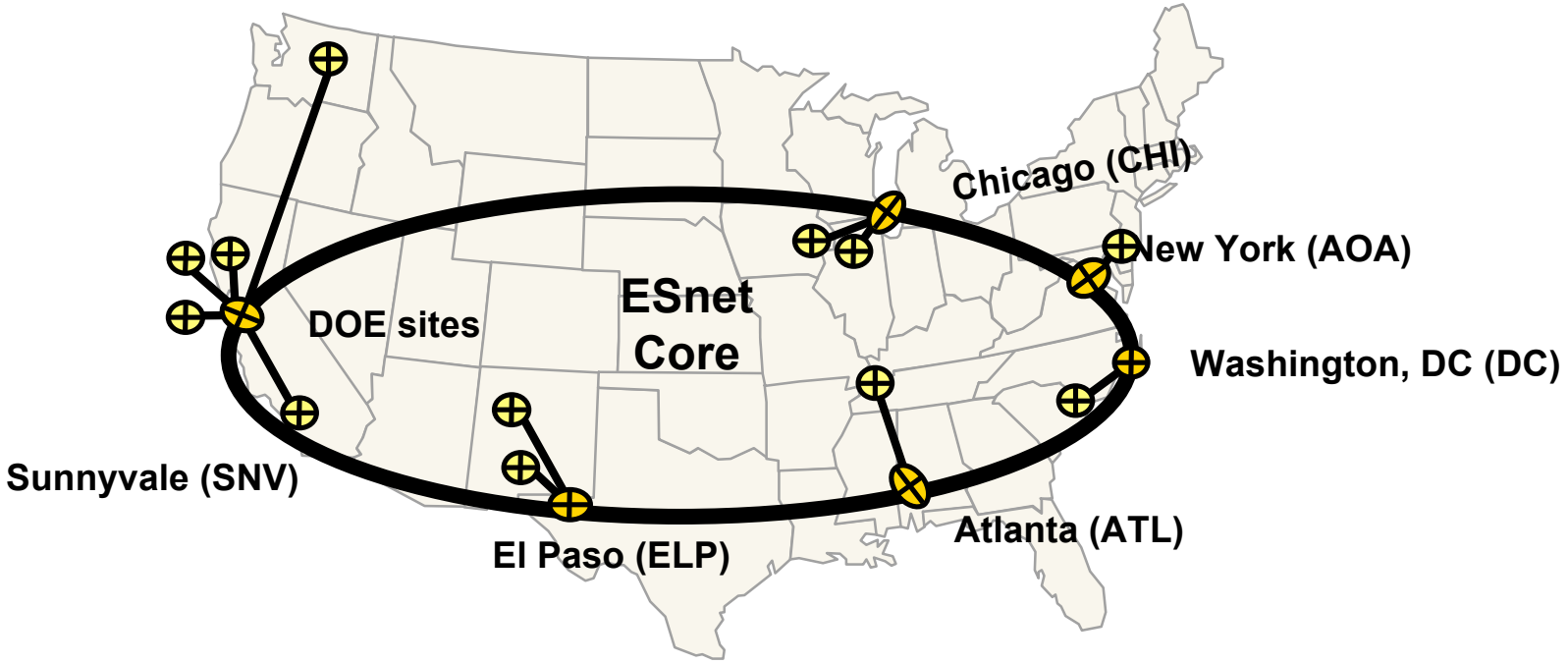
- NERSC 5L: 2005 initial delivery – 18 months earlier than base
  - 3–4 times NERSC-3 (Seaborg) in delivered performance
  - 35 Tflop/s peak, 7 Tflop/s sustained
  - Used for entire workload and has to be balanced
- Capacity Cluster: 2006–2009
  - 3 times Seaborg but for capacity work only
  - 27 Tflop/s peak, 3.7 Tflop/s sustained
  - Candidate system: blade cluster
- NERSC 6L: 2008 initial delivery – 24 months earlier than base
  - 4–5 times NERSC 5L in delivered performance
  - 167 Tflop/s peak, 40 Tflop/s sustained
  - Used for entire workload and has to be balanced

Detailed NERSC Capability Plan



New ESnet Architecture Needed to Accommodate OSC

- **The essential DOE Office of Science requirements cannot be met with the current, telecom-provided, hub-and-spoke architecture of ESnet**



The core ring has good capacity and resiliency against single-point failures, but the point-to-point tail circuits are neither reliable nor scalable to the required bandwidth

A New ESnet Architecture

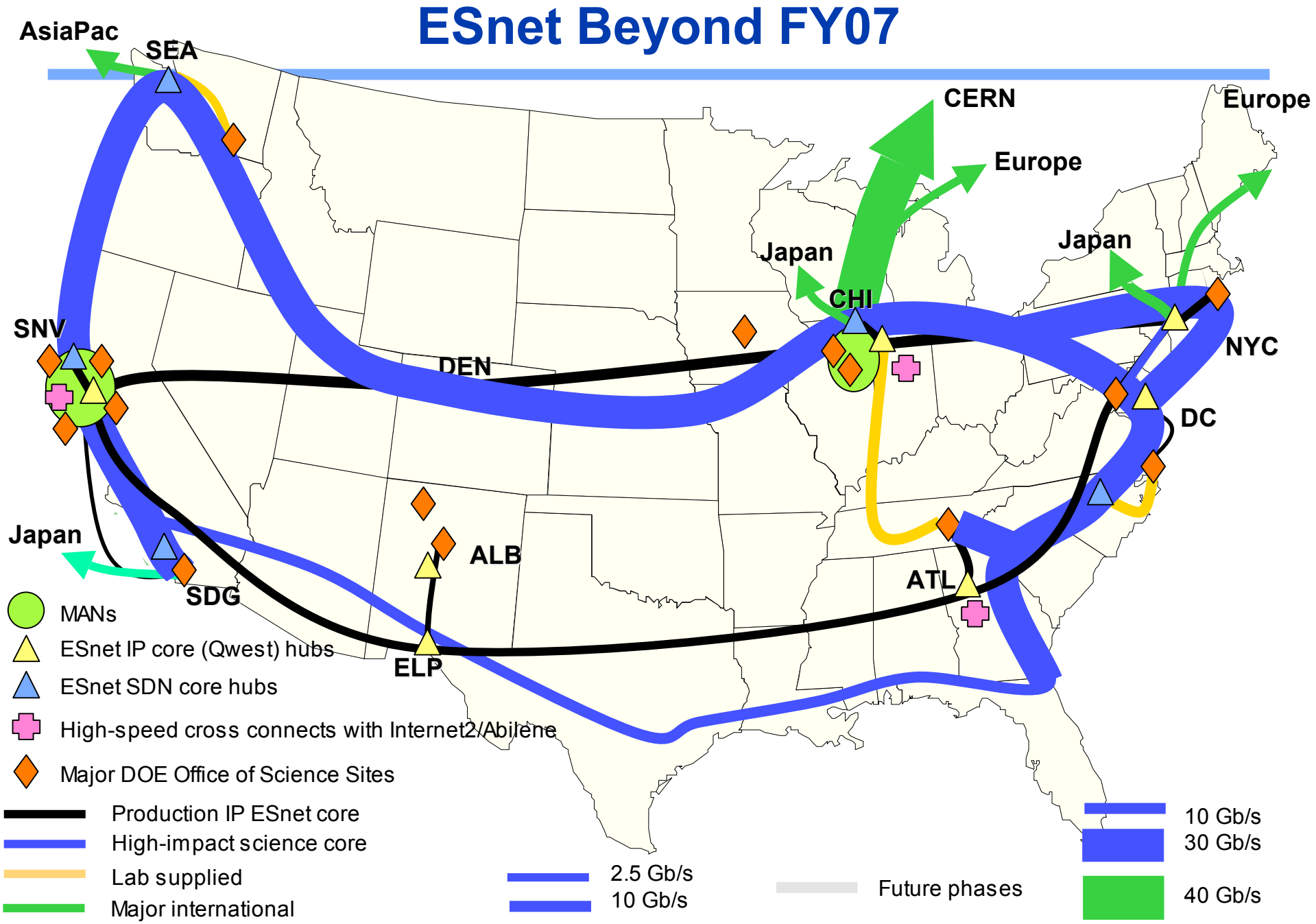
- Goals
  - Full redundant connectivity for every site
  - High-speed access for every site (at least 10 Gb/s)

- Three-part strategy

  1) MAN rings provide dual site connectivity and much higher site-to-core bandwidth

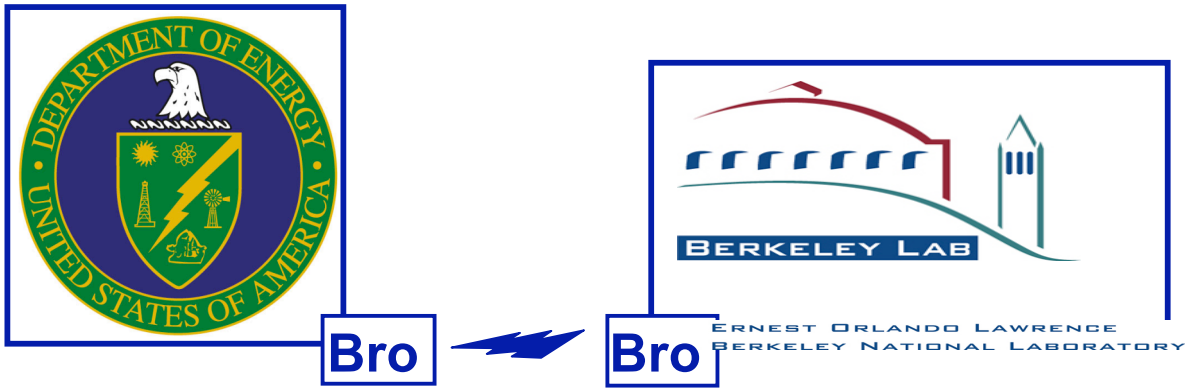
  2) A Science Data Network core for
     - Multiple connected MAN rings for protection against hub failure
     - Expanded capacity for science data
     - A platform for provisioned, guaranteed bandwidth circuits
     - Alternate path for production IP traffic
     - Carrier circuit and fiber access neutral hubs

  3) An IP core (e.g., the current ESnet core) for high reliability


ESnet Beyond FY07
AsiaPac
SEA
CERN
Europe
Europe
Japan
Japan
CHI
SNV
DEN
NYC
DC
Japan
ALB
ATL
SDG
ELP
MANs
ESnet IP core (Qwest) hubs
ESnet SDN core hubs
High-speed cross connects with Internet2/Abilene
Major DOE Office of Science Sites
Production IP ESnet core
High-impact science core
Lab supplied
Major international
2.5 Gb/s
10 Gb/s
Future phases
10 Gb/s
30 Gb/s
40 Gb/s

Cyber Security Strategy

- In 2004: DOE CIO funds Bro to be installed at HQs and other DOE sites
- Goal: National leader in cyber security research AND deployment



Summary

- NERSC has developed a path to address the increased computational needs of the Office of Science
- Leverage staff and infrastructure for rapid and cost- efficient delivery of new resources
- Partner with SciDAC projects for rapid introduction of algorithms and tools
- New ESnet architecture will provide high-speed fully redundant connectivity to all sites
- Cybersecurity research and deployment are increasingly important issues for open computing